

Kansalliskirjaston palvelut digitoijalle

Jyrki Ilva

Lassi Lager

Tieteellisen kirjallisuuden digitointipäivä 26.1.2016

Kansalliskirjastolla monta roolia

- Lakisääteisiä tehtäviä mm. kulttuuriaineistolakiin liittyen
 - Vapaakappaleet, verkkoaineistojen arkistointi
- Huolehtii omien kokoelmiensa säilytyksestä (ml. digitointi) ja aineistojen tarjoamisesta käyttäjille
- Toimii asiantuntijaorganisaationa ja tuottaa keskitettyjä kansallisia palveluita muille organisaatioille (mm. FinElib, Finna, Melinda, Finto)
- Tarjoaa organisaatioasiakkaille palveluita myös maksullisena palvelutoimintana (mm. julkaisuarkistot, Elektra, digitointipalvelut)

Vapaakappaletoiminta

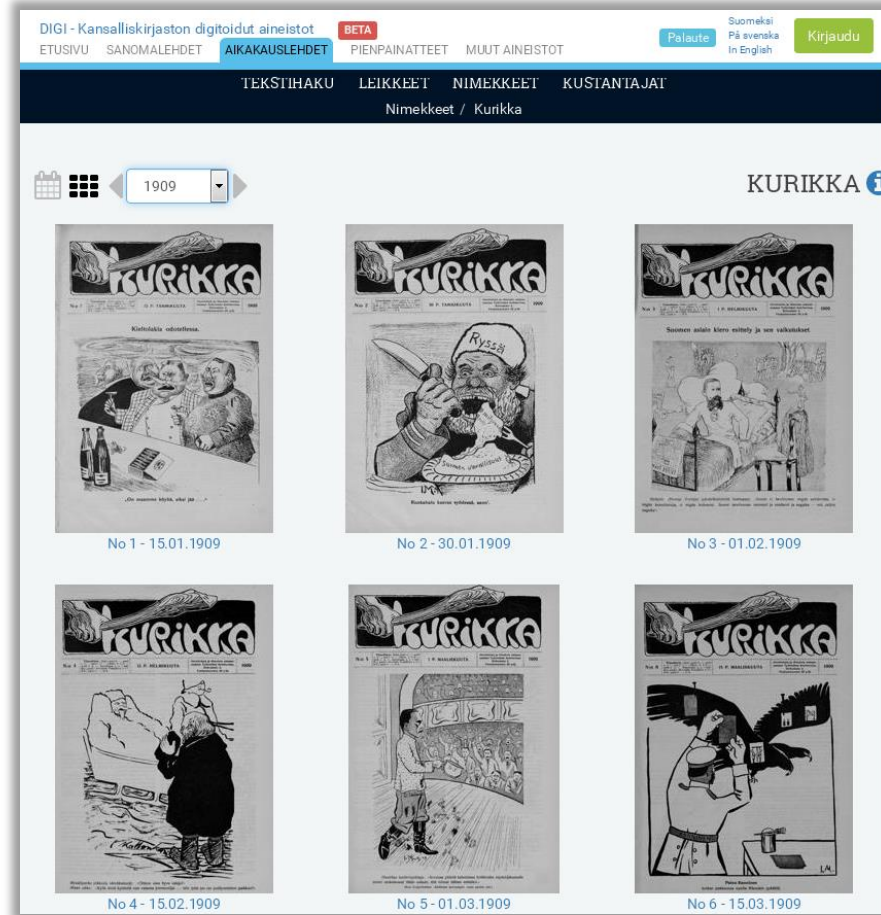
- Kansalliskirjaston ja julkaisualan toimijoiden tehtävänä on yhteistyössä tallettaa kotimainen julkaisutuotanto tutkijoiden ja muiden tarvitsijoiden käyttöön
- Tallettaminen koskee painotuotteita, äänitteitä ja muita tallenteita sekä digitaalisia julkaisuja ja muuta verkkoaineistoa
- Toiminta perustuu [lakiin kulttuuriaineistojen tallettamisesta ja säilyttämisestä](#), 1433/2007 (kulttuuriaineistolaki)
- Toiminnan tarkoituksena on kansallisen julkaisutuotannon säilyttäminen ja saattaminen tutkijoiden ja muiden tiedon tarvitsijoiden käyttöön nyt ja tulevaisuudessa
- Digitaaliset aineistot käytettävissä vapaakappalekirjastoissa ja Kansallisessa audiovisuaalisessa instituutissa

Kansalliskirjaston oma digitointi

- Kansalliskirjaston Mikkelin digitointi- ja konservointikeskukseen rakennettu 1990-luvun lopulta lähtien erityisesti massadigitointiin soveltuva tekninen infrastruktuuri
 - Tavoitteena digitointiprosessin ja digitoitujen aineistojen koko elinkaaren hallinta
 - Tavoitteena ”autenttiset ja luotettavat” digitoidut tallenteet
- Kansalliskirjaston päämääränä digitointiin ja digitaalisiin aineistoihin liittyvien kansallisten infrastruktuurien rakentaminen
 - Kokoelmien massadigitointi: ei poimita rusinoita pullasta, vaan digitoidaan laajoja yhtenäisiä aineistokokonaisuuksia
 - Keskitytty etenkin sanomalehtiin (digitoitu 1910-luvulle saakka) ja aikakauslehtiin (digitoitu pääosin vuoteen 1944)
 - Digitoitu myös kirjoja, pienpainatteita, karttoja, äänitteitä

Kansalliskirjaston digitoituja aineistoja useissa palveluissa

- [Digi – Kansalliskirjaston digitoidut aineistot](#)
 - Sanoma- ja aikakauslehtiä, pienpainatteita
 - Toimii yhteen digitointikeskuksen prosessien kanssa
- [Doria](#)
 - Kirjoja, pienpainatteita, karttoja, musiikkia
 - Osana julkaisuarkistopalvelua, jossa muidenkin organisaatioiden aineistoja
- [Fragmenta membranea](#)
 - Keskiaikaisia pergamenttejä
- [Fenno-ugrica](#)
 - Suomalais-ugrilaisilla kielillä julkaistuja lehtiä ja kirjoja



Digitointihankkeita ja yhteistyöprojekteja

- [Aviisi](#) – tekijänoikeuden alaisia sanomalehtiä verkkoon yhteistyössä kustantajien ja Kopioston kanssa
- [Klassikkokirjasto](#) – tekijänoikeudesta vapaata kotimaista kaunokirjallisuutta
- [Fenno-ugrica](#) - Venäjällä digitoituja Neuvostoliiton aikaisia suomalais-ugrilaisilla kielillä julkaistuja lehtiä ja kirjoja
- [Turun akatemian väitöskirjat 1640-1828](#) – digitoitu Ilkka ja Ulla Paateron rahaston varoilla
- Suomi 100 (2014-2017): Suomen historiaan ja kulttuuriin liittyvää aineistoa
- Vuonna 2010 suunniteltu Tieteellisten lehtien digitointihanke ei sen sijaan saanut rahoitusta

Kansalliskirjaston digitointipalvelut

- Korkeatasoisia digitointipalveluita myös ulkopuolisille asiakkaille
- Lisätietoa:
Kansalliskirjaston digitointi- ja konservointikeskus, Mikkeli
kk-dimiko-asiakaspalvelu@helsinki.fi
<http://www.kansalliskirjasto.fi/fi/kirjastoala/dimiko.html>

Kansalliskirjaston digitointikeskuksen palveluita kannattaa hyödyntää, jos

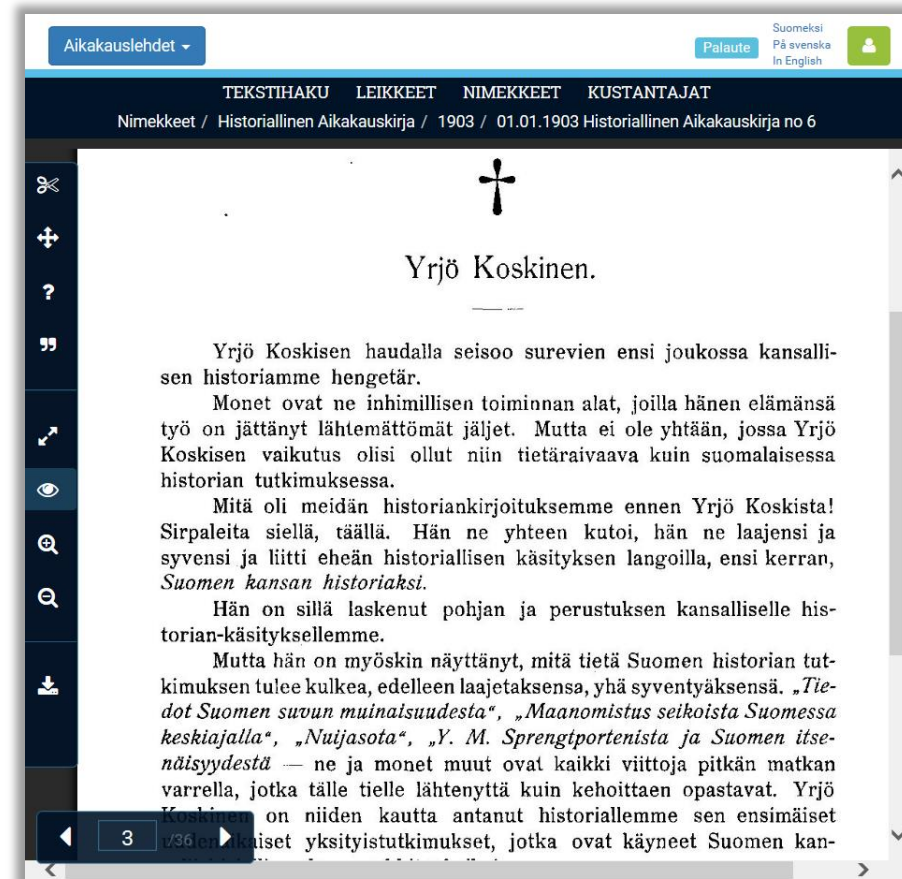
- originaaliaineistona käytetään Kansalliskirjaston kokoelmiin kuuluvaa materiaalia
- digitoitavaa aineistoa on paljon, esim. laajoja julkaisukokonaisuuksia
- aineisto on huonokuntoista, herkästi vaurioituvaa, suurikokoista tai sidottua
- aineistotyyppi on mikrofilmi, negatiivi tai dia
- halutaan korkealaatuista, esimerkiksi julkaisutoimintaan riittävää skannauslaatua
- halutaan tallentaa digitointiprosessin tekniset vaiheet (käytetyt laitteet, laatu, käsittelyhistoria)
- aineiston käsittelylle halutaan varmistaa turvalliset, arkistovaatimukset täyttävät olosuhteet
- tilaajalla on tarvetta skannatun aineiston jälkikäsittelylle
- tilaaja tarvitsee digitoinnin lisäksi aiheeseen liittyvää suunnittelua tai konsultointia
- *halutaan varmistaa digitoidun aineiston pitkäaikaissäilytyskelpoisuus*

Digitoinnin vaiheita

1. Aineiston valinta, suunnitelmat, resurssit, alustan valinta jne
2. Skannaus
3. Kuvankäsittely
4. Tekstintunnistus (OCR)
 - Vapaasanahaku tekstisisältöön
 - Oikoluvun jälkeen sisältö käytettävissä tekstimuodossa
 - Ohjelmistoja esim. Abbyy Finereader, Adobe Acrobat
 - Huono painolaatu, vanhat kirjasintyypit ja dokumenttien kieli voivat aiheuttaa haasteita
5. Rakenteistaminen
6. Lopputuote ja pitkäaikaissäilytys
7. Käyttöön saattaminen

Mitä prosessissa syntyy?

- Master-kuvat häviöttömästi pakattuna
- Digitoidulle objektille pysyvä tunniste ja linkitys luettelointitietoon
- Käyttöoikeustiedot
- Käyttöversiot
- Digitoidun aineiston syntyhistoria
 - Milloin digitoitu ja millä laitteella?
 - Aineistolle tehty käsittely?



Digitoinnin lopputuote: METS-paketti

- Kansalliskirjaston omista digitoinneista tuotetaan pitkäaikaissäilytyskelpoinen (KDK-PAS-yhteensopiva) METS-paketti, joka sisältää
 - masterkuvat häviöttömästi pakattuna
 - käyttökuvat JPEG 200dpi
 - thumbnail JPEG
 - tunnistettu teksti ALTO xml muodossa
 - nide PDF 200dpi, sisältäen tunnistetun tekstin
 - METS xml - objektin rakenne, kuvaileva-, tekninen ja hallinnollinen metatieto
- METS = [Metadata Encoding and Transmission Standard](#)

[Etusivu](#)[Ajankohtaista](#)[Tietoa hankkeesta](#)[Asiakasliittymä Finna](#)[Pitkäaikaissäilytys](#)[Toteuttamis-
suunnitelma](#)[Määrittely ja
dokumentit](#)[Yhteystiedot](#)[Kokonaisarkkitehtuuri](#)[Digitointi](#)[Yhteistyö](#)[Aineistot](#)[Kansalliset ja EU:n
linjaukset](#)[Europeana](#)

KDK:n hallinnolliset ja rakenteelliset metatiedot ja aineiston paketointi

KDK-hankkeessa on määritelty pitkäaikaissäilytyksen vaatimat hallinnolliset ja rakenteelliset metatiedot sekä ohjeistus aineiston paketointiin.

Määritys kertoo mitä pakollisia ja suositeltavia metatietoja on aineistoihin liitettävä sekä miten digitaaliset objektit ja niihin liittyvät metatiedot tulee paketoita KDK:n METS-profiilia hyödyntäen.

Määritystä päivitetään pitkäaikaissäilytyksen suunnittelun edetessä.

[KDK:n hallinnolliset ja rakenteelliset metatiedot ja aineiston paketointi 1.4 \(PDF, 14.4.2014\)](#)

Metatiedot

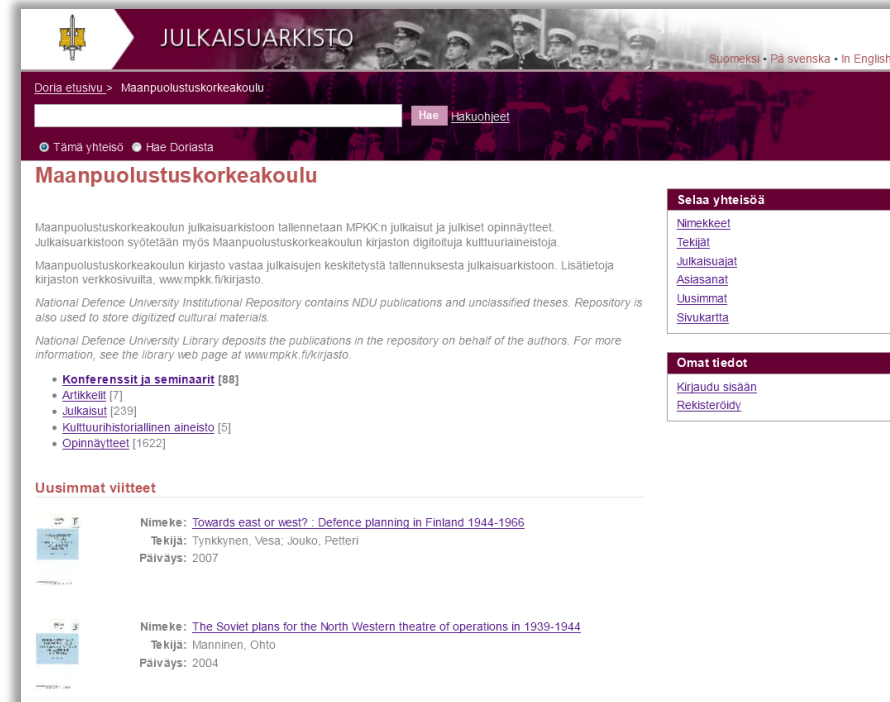
- **Kuvaileva metatieto** kuvaa aineiston (esim. digitoidun kirjan tai artikkelin) sisältöä ja luonnetta
 - Kuvailevan metatiedon sisällön tuottamisessa on suositeltavaa käyttää **standardinmukaisia metatietoformaatteja** (esim. MARC 21 tai Dublin Core), **valmiita sanastoja**, luokituksia ja ontologioita (esim. YSA/YSO Fintossa) ja **yhdessä sovittuja käytänteitä ja soveltamisohjeita** (esim. ARTO-OJS-yhteensopivuus)
- **Hallinnollinen metatieto** määrittelee aineiston hallintaan tarvittavat hallinnolliset ja tekniset tiedot, esimerkiksi aineiston käyttöehdot ja pitkäaikaissäilytykseen tarvittavat tekniset tiedot
- **Rakenteellinen metatieto** kuvaa aineiston rakennetta ja sen järjestystä, esimerkiksi aineiston eri osien suhdetta toisiinsa (esim. digitoitu lehti tai kirja ja sen artikkelit)

Digitoitu aineisto käyttöön

- Alustan valinnassa kannattaa miettiä aineiston pitkäaikaista käyttöä (tuleva ylläpito, versiopäivitykset, jne.)
 - Kannattaako rakentaa itse vai hyödyntää jotain keskitettyä infrastruktuuria?
 - Myös aineiston luonteella ja käytön laajuudella merkitystä: Tuleeko vapaaseen käyttöön? Tarvitaanko työkaluja käyttöoikeuksien rajoittamiseen / käytön seurantaan?
- Tekijänoikeus- ja tietosuojakysymykset?
- Standardinmukaiset ja riittävät metatiedot
- Pysyvät tunnisteet, esim. URN tai DOI
- Metatiedot kirjastojen tietokantoihin?
- Avoimet hakurajapinnat?
- Hakukoneille (esim. Google) pääsy metatietoihin ja tiedostoihin?

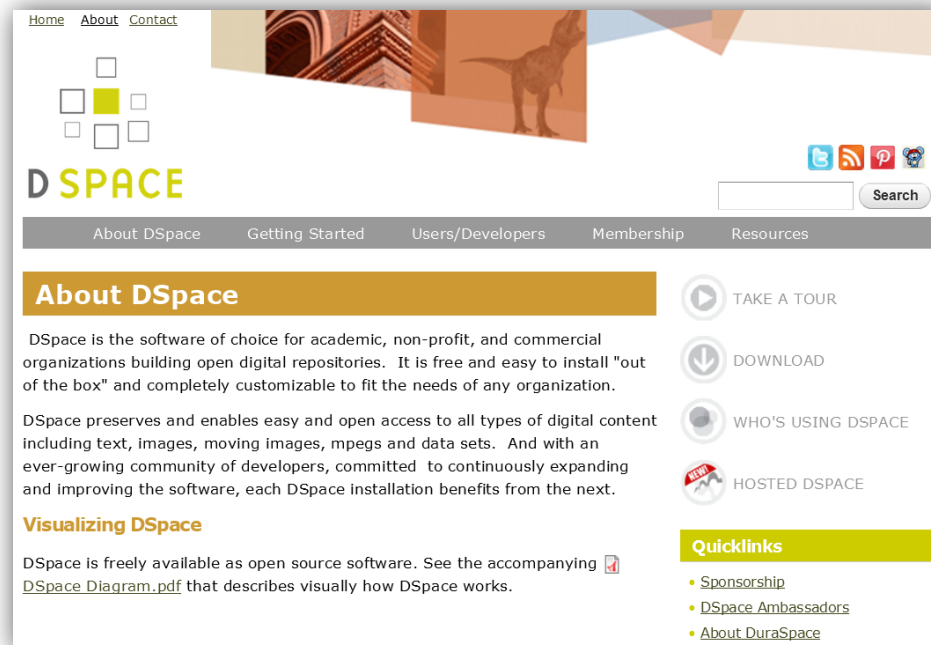
Kansalliskirjaston palvelut: Julkaisuarkistot

- Kansalliskirjasto tarjoaa julkaisuarkistokäyttöön soveltuvaa infrastruktuuria asiakasorganisaatioille
- Maksullista palvelutoimintaa
 - Palvelun ylläpito ja kehittäminen rahoitetaan vuosimaksuilla
 - Tällä hetkellä mukana 46 organisaatiota + Kansalliskirjaston omia aineistoja
 - Yliopistoja, ammattikorkeakouluja, tutkimuslaitoksia, virastoja,...
 - Asiakasorganisaatioilla oma kokoelma ja alakokoelmia joko omassa tai jaetussa palvelussa



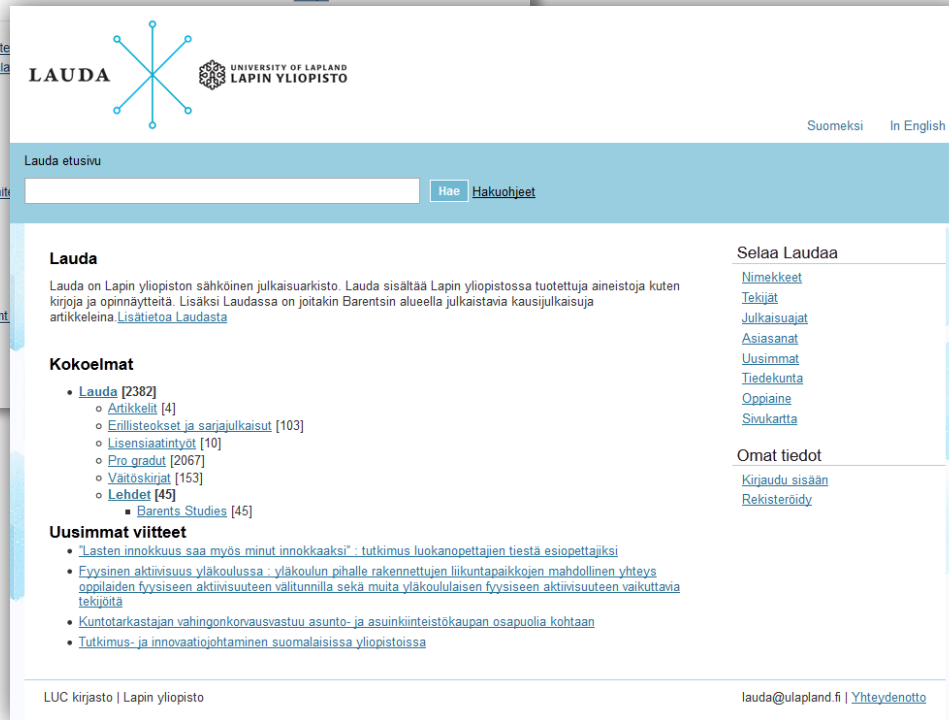
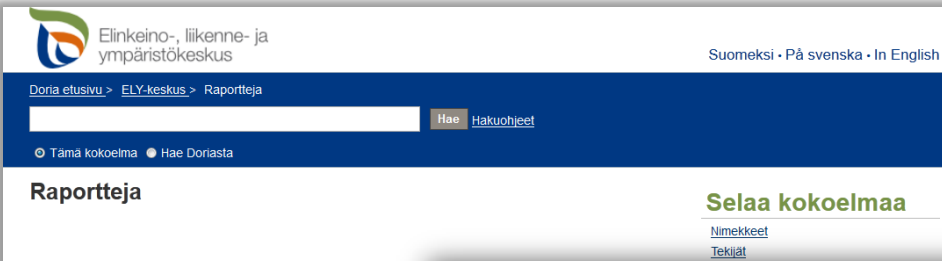
Kansalliskirjaston julkaisuarkistopalvelut

- Palvelut perustuvat avoimen lähdekoodin DSpace-ohjelmistoon
 - Tukee monenlaisia tiedostoformaatteja
 - Metadata Dublin Core -formaattissa
 - Dokumentteja mahdollista syöttää yksitellen tai eräajoina
 - Vahvuuksia: kansainvälinen kehittäjäyhteisö, tuki syöttöprosesseille, kokotekstihaku, tekniset rajapinnat, aineistot indeksoituvat hyvin myös Googleen



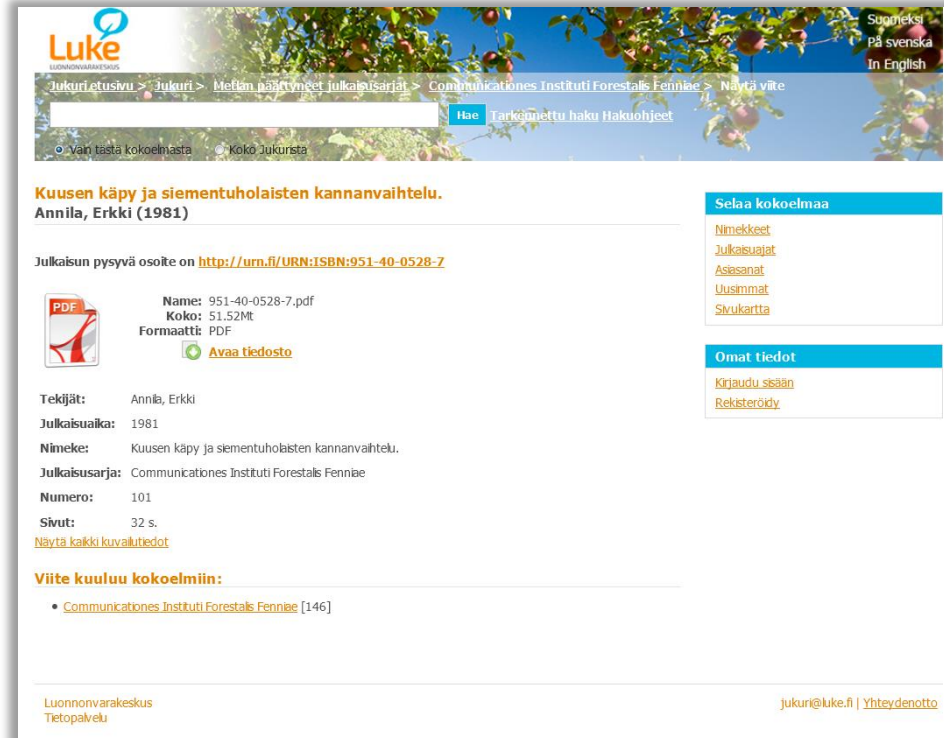
Useita erillisiä DSpace-palveluita

- [Doria](#)
- [Fenno-ugrica](#)
- [Fragmenta
membranea](#)
- [Jukuri](#)
- [Julkari](#)
- [Lauda](#)
- [Tampub](#)
- [Theseus](#)
- [Varia](#)



Julkaisuarkistopalvelun sisällöt

- Kansalliskirjasto vastaa palvelun teknisestä ylläpidosta ja kehittämisestä; asiakasorganisaatiot vastaavat kokoelmistaan ja aineistoistaan
- Käytetään asiakasorganisaatioissa etenkin opinnäytteiden, tieteellisten julkaisujen ja sarjajulkaisujen tallentamiseen ja julkaisemiseen
- Myös digitoituja aineistoja, yleensä pdf-formaatissa



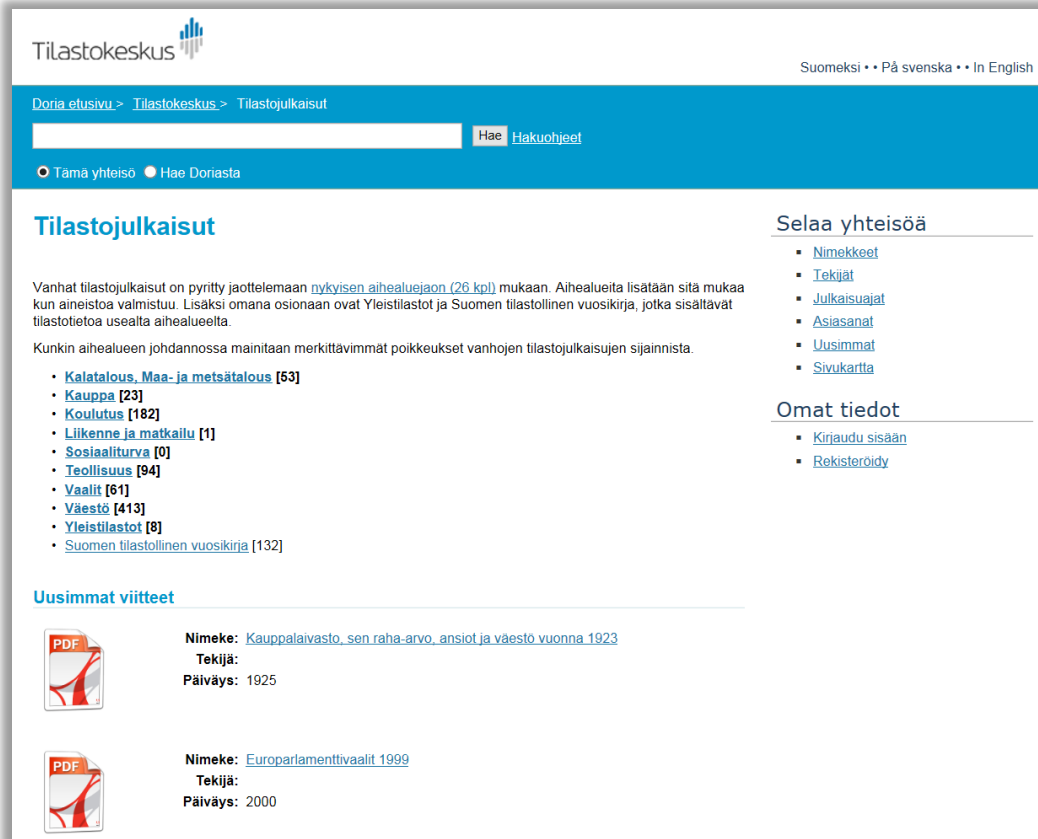
Esim. Yhteiskuntapolitiikka ja Alkoholipolitiikka

- Yhteiskuntapolitiikka- ja Alkoholipolitiikka-lehtien takautuvien vuosikertojen artikkelit THL:n julkaisuarkistoon Julkarissa
- Hyödynnettiin Artosta poimittuja artikkelien kuvailutietoja
- Digitoidut Alkoholipolitiikat pilkottiin artikkeleiksi
- Tiedostot ja tietueet yhdistettiin toisiinsa ja ajettiin eräajona Julkariin



Esim. Tilastokeskuksen tilastojulkaisut

- Noin tuhat Tilastokeskuksen digitoimaa tilastojulkaisua sijoitettu Doriaan
- Tilastokeskuksella oma päätason kokoelma jonka alla alakokoelmia
- Kuvailutiedot ja tiedostot siirtyvät rajapinnan kautta Tilastokeskuksen järjestelmästä Doriaan



Tilastokeskus

Suomeksi • På svenska • In English

[Doria etusivu](#) > [Tilastokeskus](#) > [Tilastojulkaisut](#)

[Hae](#) [Hakuohjeet](#)

☐ Tämä yhteisö ☒ Hae Doriasta

Tilastojulkaisut

Vanhhat tilastojulkaisut on pyritty jaottelemaan [nykyisen aihealuejaon \(26 kpl\)](#) mukaan. Aihealueita lisätään sitä mukaa kun aineistoa valmistuu. Lisäksi omalla osionaan ovat Yleistilastot ja Suomen tilastollinen vuosikirja, jotka sisältävät tilastotietoa usealta aihealueelta.

Kunkin aihealueen johdannossa mainitaan merkittävimmät poikkeukset vanhojen tilastojulkaisujen sijainnista.

- [Kalatalous, Maa- ja metsätalous](#) [53]
- [Kauppa](#) [23]
- [Koulutus](#) [182]
- [Liikenne ja matkailu](#) [1]
- [Sosiaaliturva](#) [0]
- [Teollisuus](#) [94]
- [Vaalit](#) [61]
- [Väestö](#) [413]
- [Yleistilastot](#) [8]
- [Suomen tilastollinen vuosikirja](#) [132]

Selaa yhteisöä

- [Nimekkeet](#)
- [Tekijät](#)
- [Julkaisuajat](#)
- [Asiasanat](#)
- [Uusimmat](#)
- [Sivukartta](#)

Omat tiedot

- [Kirjautu sisään](#)
- [Rekisteröidy](#)

Uusimmat viitteet



Nimeke: [Kauppalaivasto_sen raha-arvo_ansiot ja väestö vuonna 1923](#)
Tekijä:
Päiväys: 1925



Nimeke: [Europarlamenttivaalit 1999](#)
Tekijä:
Päiväys: 2000

Maksullisuus vs. avoimuus?

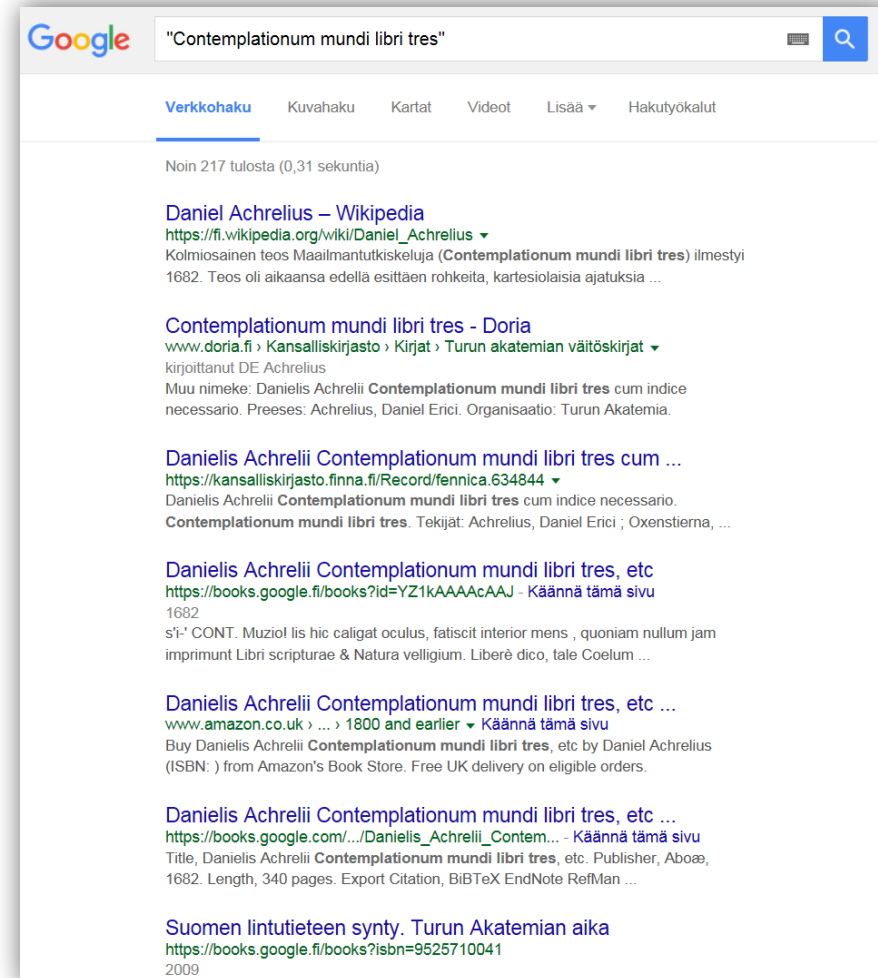
- Kirjastot hankkivat maksullisia e-aineistoja yli 20 miljoonalla eurolla vuodessa
 - Suurin osa etenkin tieteellisten kirjastojen hankinnoista tehdään Kansalliskirjastossa toimivan FinELibin kautta
 - Hankinnat kohdistuvat lähinnä suuriin aineistokokonaisuuksiin
 - Palveluiden laadun ja käyttöehtojen suhteen melko tiukat vaatimukset
- Etenkin tieteellisen julkaisemisen osalta pyritään vähitellen siirtymään open access -julkaisemiseen
 - Palvelee tutkimusta ja tiedon saatavuutta
 - Kv. julkaisutoimintaan liittyy merkittäviä taloudellisia intressejä
 - Yhteishanke TSV:n kanssa: Kotimaiset lehdet avoimiksi ja vaikuttamaan

Kansalliskirjaston palvelut: Elektra ja Peri+

- [Elektra](#) (yhteistyössä Kopioston kanssa)
 - Tällä hetkellä aktiivisesti mukana n. 35 tieteellistä lehteä
 - Lehdet pilkottu artikkeleiksi, yli 34.000 artikkelia
 - Suurin osa maksullista aineistoa, joka käytössä mm. monissa yliopistoissa, kaikissa amk:eissa ja yleisissä kirjastoissa
 - Aineisto digitaalisessa muodossa kustantajilta; hanke alkoi 1996, mutta mukana myös joitakin takautuvasti digitoituja lehtiä
 - Kuvailutiedot löydettävissä mm. Arton, Dorian, Finnan ja Googlen kautta
- Peri+ (yhteistyössä Kopioston kanssa)
 - Mikkelissä digitoituja tieteellisten ja kulttuurilehtien takautuvia vuosikertoja
 - Sijaitsevat Mikkelin Digissä; kuuluvat Elektra-sopimukseen

Aineistojen näkyvyys ja löydettävyys (1)

- Google ja muut verkon yleiset hakukoneet merkittäviä
 - Haravointi ja kokotekstiindeksointi
- Monissa palveluissa valtaosa kävijöistä tulee Googlen kautta
- Aineistojen pitää kuitenkin ensin päätyä Google-haun piiriin



Aineistojen näkyvyys ja löydettävyys (2)

- Tieteelliset aineistot ja viittaustietokannat
 - Sisältävät viitetietojen lisäksi tietoja myös julkaisuihin sisältyvistä ja niihin kohdistuvista viittauksista
 - Google Scholar vapaassa käytössä; haravoi aineistoja verkosta
 - Maksulliset Web of Science (Thomson Reuters) ja Scopus (Elsevier) valikoivat indeksoimansa lehdet tarkemmin
- Kotimaiset viitetietokannat - Fennica, Melinda, Arto
 - Kirjojen viitetiedot Fennicaan/Melindaan?
 - Artikkelimetatiedot ARTOon?
 - Fennicasta ja ARTOsta tiedot menevät Finnaan

Artikkelimetatiedot ARTOon?

<https://wiki.helsinki.fi/pages/viewpage.action?pageId=112442496>

- Tallennus Artiva-lomakkeella, haravointi ARTOon tai valmiiden metatietojen hyödyntäminen ARTOssa?
- ARTOsta mahdollisuus siirtää viitetietoja myös muihin järjestelmiin ja formaatteihin
- ARTOn tiedot indeksoidaan Finnan kautta Googleen, tieteellisten artikkeliviitteiden Google Scholar -haravoinnista neuvoteltu

Arto

Kuvailun kohde ☒ Lehtiartikkeli ☐ Artikkeli kokoc

Lehden nimi

Vuosi Vol. Nro.

Artikkelin otsikko

Linkki kokotekstiin

Tekijän sukunimi **Etunimi**

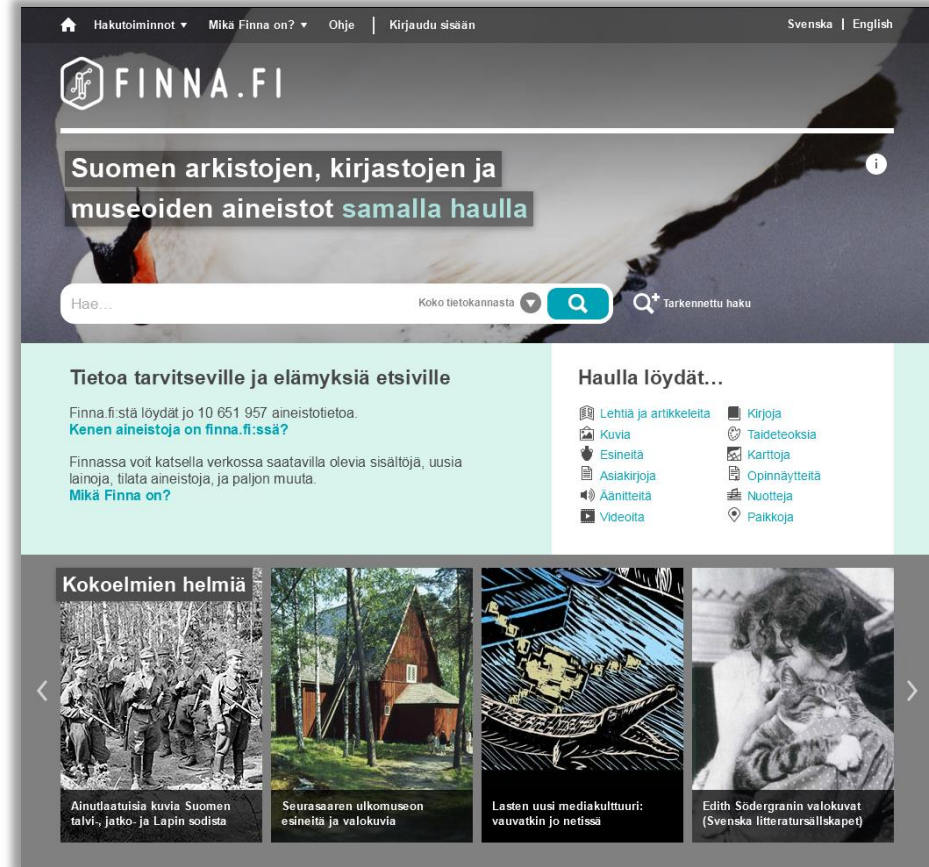
Organisaatio

Tiivistelmä/abstrakti

Artikkelin tyyppi

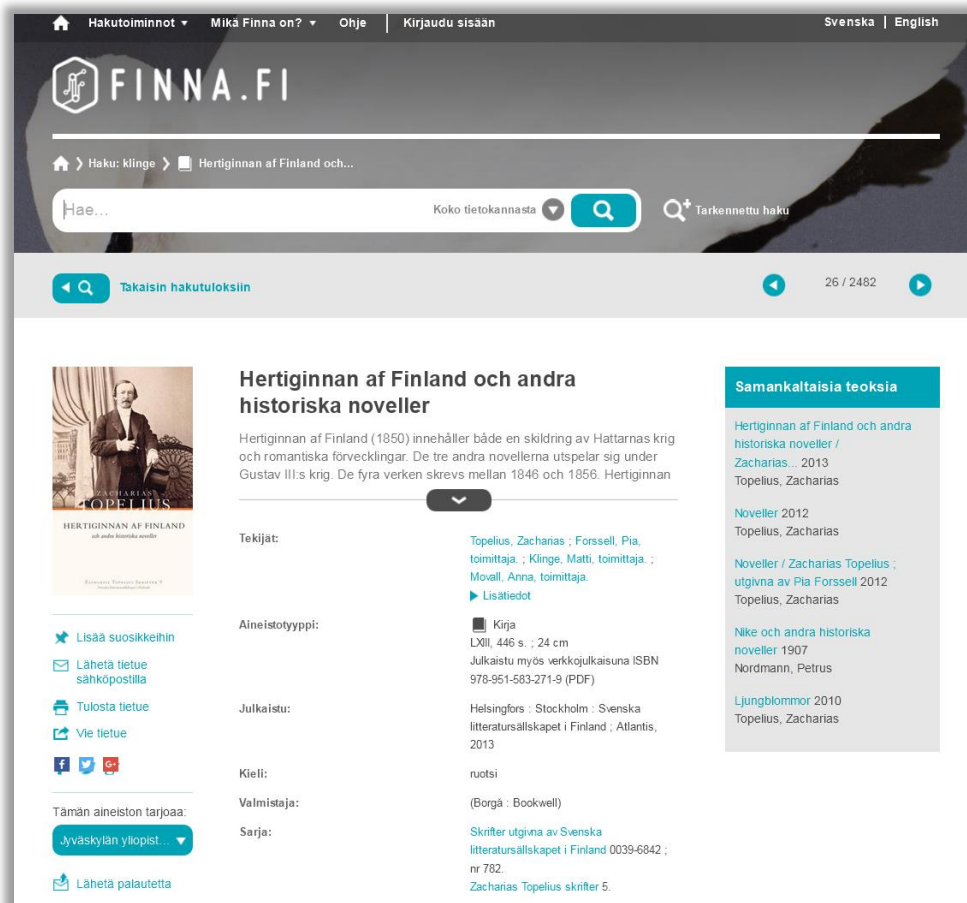
Finna-palvelu

- Finna on kansallisen digitaalisen kirjaston asiakasliittymä, <https://www.finna.fi/>
- Kansallinen näkymä ja organisaatiokohtaisia näkymiä
- Haravoi aineistojen metatiedot taustajärjestelmistä
- Korvaa kirjastojen aiemmat hakukäyttöliittymät
- Kuvailutiedot Finnan kautta Googlen haravoitavissa; hyödynnettävissä myös avoimen rajapinnan (API) kautta



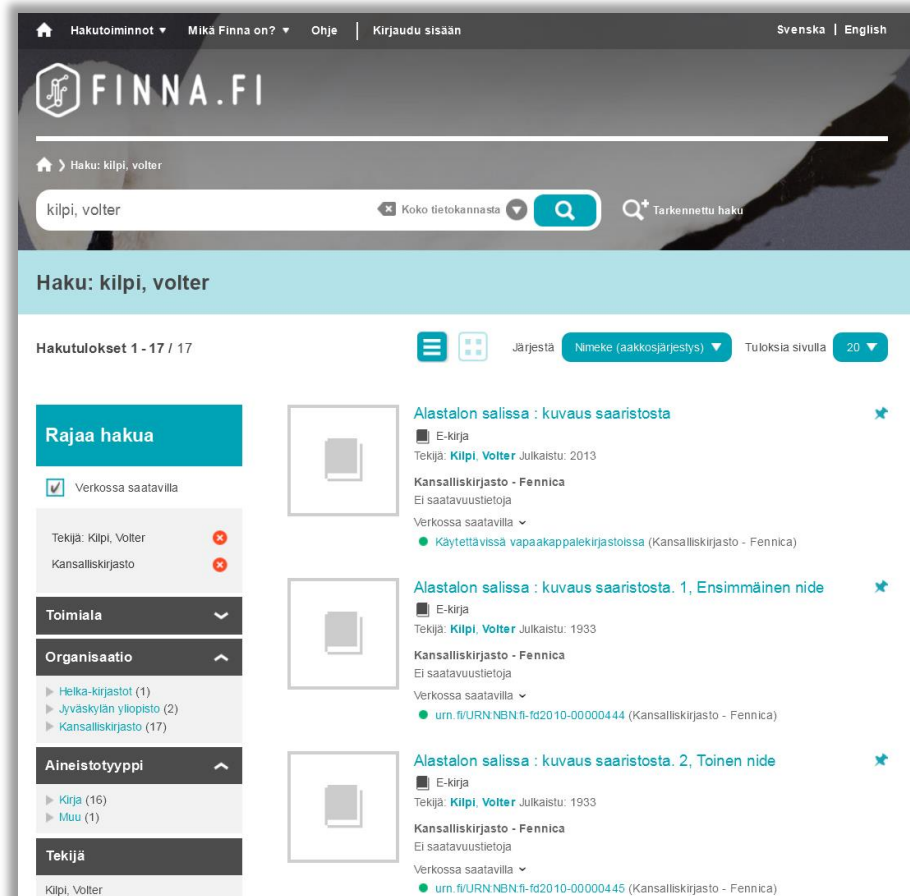
Aineistojen näkyvyys Finnassa

- Kuvailutiedot haravoidaan Finnaan pääasiassa asiakaskirjastojen, -arkistojen ja -museoiden järjestelmistä, mutta myös Fennicasta ja ARTOsta
- **Fennicaan** (kansallisbibliografia) tallennetaan Suomessa julkaistavien kirjojen tiedot
 - Tieteellisten kirjojen tiedot menevät usein myös muuta kautta, eli mukana olevien kirjastojen tietokantojen kautta Finnaan



Digitoitujen teosten tiedot?

- Kansalliskirjasto kerää ja vie digitaalisten julkaisujen metatiedot mahdollisuuksien mukaan Fennicaan
 - Kansalliskirjastossa suunnitellaan kotimaisten julkaisuarkistojen keskitettyä haravointia vapaakappalepalvelimelle
 - Finna-asiakkaiden järjestelmistä metatiedot voidaan haravoida myös suoraan Finnaan
 - Edellyttää OAI-PMH-haravointimahdollisuutta
 - Vapaakappaleina haravoitujen aineistojen metatiedot mahdollisesti Fennican kautta Finnaan



Pysyvät tunnisteet ja osoitteet

- Pysyvä tunniste, persistent identifier, PID
 - Ainutkertainen, viittaa johonkin tiettyyn dokumenttiin/entiteettiin tai sen versioon
- Pysyvien tunnisteiden pohjalta on mahdollista muodostaa myös pysyviä verkko-osoitteita
 - Edellyttää sitä että dokumentin sijainti resolvointipalvelun (esim. urn.fi) tiedossa; osoitteet vaativat ylläpitoa
- Dokumenttien tunnistaminen
 - Tunnistejärjestelmiä: URN, DOI, Handle, ARK
 - Kirjoille ja vastaaville ISBN, lehdille (jatkuville julkaisuille) ISSN
- Henkilöiden ja yhteisöjen tunnistaminen
 - Tunnistejärjestelmiä: ISNI, ORCID

DOI ja URN

- DOI (Digital Object Identifier) vakiintunut tieteellisten artikkelien de facto standarditunnisteeksi
 - Tunnukset maksullisia; tunnuksia myöntäviä organisaatioita useita, tärkein CrossRef
 - Tieteellisen julkaisemisen kansainvälisiä infrastruktuureja rakennettu DOI:n ympärille
- URN (Uniform Resource Name)
 - Verkkojulkaisun ainutkertainen ja pysyvä tunniste
 - Tunnukset maksuttomia, käytössä laajasti mm. Suomessa

Kansalliskirjaston palvelut: URN

- Kansalliskirjaston URN-rekisteri toimii hajautettuna palveluna (toistaiseksi) suomalaisten yliopistojen ja muiden pysyväluonteisten organisaatioiden kesken
- Tällä hetkellä Kansalliskirjasto ei tarjoa URN-palveluita pienille kustantajille, julkaisijoille tai yksityishenkilöille
- URN-tunnisteet käytössä esim. Kansalliskirjaston ylläpitämissä julkaisuarkistoissa
- URN-palvelua kehitetään osana ATT-hanketta
- Lisätietoa: kk-urn@helsinki.fi

Kansalliskirjasto jakaa myös ISSN- ja ISBN-tunnuksia

- Jos tieteellistä lehteä tai sarjaa aletaan julkaista (myös) verkossa, verkkoversiolle tulisi pyytää oma ISSN
- Vastaavasti painetun kirjan PDF-versiolle oma ISBN
- Verkkojulkaisulla voi olla sekä ISBN että URN – silloin URN on yleensä muodostettu ISBN:n pohjalta.

Lisätietoa: <http://www.kansalliskirjasto.fi/fi/julkaisuala>

Tekijöiden tunnistaminen: ISNI ja ORCID

- [ISNI](#) on henkilöiden ja yhteisöjen nimien eli ”julkisten identiteettien” yleinen kansainvälinen standarditunniste
 - Taustavoimina etenkin tekijänoikeusjärjestöt ja kirjastot
 - Suomessa kansallinen ISNI-keskus toimii Kansalliskirjastossa
 - Perustuu henkilö- ja julkaisutietojen pohjalta keskitetysti tehtyyn tunnistamiseen
- [ORCID](#) suunniteltu tieteellisen julkaisemisen tarpeisiin
 - Syntynyt suurten tieteellisten kustantajien aloitteesta, mukana myös järjestelmätoimittajia ja tutkimushallinnon organisaatioita
 - CSC koordinoi käyttöönottoa suomalaisissa tutkimusorganisaatioissa
 - Tutkija hallinnoi itse tietojaan; tunniste kulkee julkaisujen mukana

Pitkääaikaissäilytys

- Tarkoituksena säilyttää digitaaliset aineistot käyttökelpoisina useiden vuosikymmenien tai satojen vuosien ajan
- Mahdollistaa digitaalisten aineistojen hyödyntäminen tuleville sukupolville
- Toteutuksen näkökulmasta korostaa jatkuvuussuunnittelua: toiminnan elinkaari ylittää kaikkien toteutuksen osakokonaisuuksien käyttöajan
- Bittien säilyttäminen melko yksinkertaista (mutta ei riitä)
 - esim. saman aineiston vähintään kolmen eri kopion tarkistussummia verrataan keskenään. Jos yksi kopio poikkeaa muista, se tulkitaan vioittuneeksi ja korvataan ehjällä kopiolla

Säilyttämisen vaatimustasot

Lähde: CSC

Alkuperäisen käyttökokemuksen säilyttäminen

- Vaativin, mutta käyttäjän kannalta usein mielenkiintoisin taso
- Esim. tietokonepelin pelaaminen samoin kuin 30 vuotta sitten

Sisällön ymmärrettävyyden säilyttäminen

- Teksti, kuvat ja muu olennainen sisältö voidaan tulkita
- Ymmärrettävyys riippuu myös kohdeyleisöstä
- Käytön ja uuden tiedon tuottamisen kannalta yleensä tärkein taso

Bittien säilyttäminen

- Varmistetaan alkuperäisten ykkösten ja nollien säilyminen ja luettavuus, sisällön tulkinnasta ei takeita
- Säilyttämisen perusta ja minimitaso säilytyksen aloittamiseen: Suunnitelmallisesti tehtynä vaatimustasoa voidaan nostaa myöhemmin

Säilyttämisen menetelmiä

Lähde: CSC

Alkuperäisen käyttökokemuksen säilyttäminen

- **Emulaatio**
 - Emulaattoreiden kehittäminen uusille alustoille, formaattien seuranta jne.
 - Aktiivinen testaaminen ja seuranta

Sisällön ymmärrettävyyden säilyttäminen

- **Migraatio**
 - Tiedostomuotojen kehityksen seuranta (formaattikirjasto)
 - Muunnosten kehittäminen, testaaminen, toteuttaminen ja valvonta
 - Varautuminen palautuksiin

Bittien säilyttäminen

- **Eheyden varmistaminen**
 - Tiedostojen validointi ja valvonta
 - Kopioiden hallinta
 - Koskee sekä objekteja että metadataa

Kansalliset pitkäaikaissäilytysratkaisut

- **KDK-PAS**

(<http://www.kdk.fi/index.php/fi/pitkaaikaissailytys>)

- Osa Kansallinen digitaalinen kirjasto- eli KDK-hanketta
- Digitaalisten kulttuuriperintöaineistojen keskitetty pitkäaikaissäilytyspalvelu
- Tarkoitettu mm. Kansalliskirjaston keräämää kotimaista julkaisutuotantoa varten (ml. Verkoarkisto)
- Tuotantokäytössä 11/2015-

- **Tutkimus-PAS** (<http://avointiede.fi/tutkimus-pas>)

- Osa Avoin tiede ja tutkimus- eli ATT-hanketta
- Tutkimusten tulosten pitkäaikaissaatavuus
- Tulossa

- Molemmista OKM:n rahoittamista PAS-palveluista vastaa CSC

Digitoidun aineiston pitkäaikaissaatavuuden takaaminen?

Mahdollisuuksia:

1. Digitaalisena kulttuuriperintöaineistona KDK-PASiin?
 2. Tutkimusaineistona Tutkimus-PASiin?
 3. Muu jaettu PAS-ratkaisu, esim. CLOCKKS (<https://www.clockss.org/>)
 4. Oma pitkäaikaissäilytysratkaisu (yleistä esim. Yhdysvalloissa)
- Kansalliskirjasto saanut toistaiseksi OKM:ltä myönnön vain verkkoarkiston siirtämiseksi KDK-PAS-palveluun (siirrot alkoivat 11/2015)
 - Muulla tavoin saaduille digitaalisille vapaakappaleille ei ole vielä KDK-PAS-mandaattia
 - Kansalliskirjaston omillekaan digitoiduille kulttuuriperintöaineistoille ei ole vielä KDK-PAS-mandaattia.
 - Kansalliskirjaston digitoidut aineistot ja suuri osa luovutuksina saaduista digitaalisista julkaisuista kuitenkin ”PAS-kelpoisia”

TAJUA-projekti ja pitkäaikaissaatavuus

- Kansalliskirjasto selvittää kotimaisten tieteellisten julkaisujen pitkäaikaissaatavuutta tukevia ratkaisuja osana OKM:n Avoin tiede ja tutkimus -hankkeen rahoittamaa [TAJUA-projektia](#) (2015-2017)

Kiitos!

jyrki.ilva@helsinki.fi
lassi.lager@helsinki.fi